Vocabularies and Common Data Elements (VCDE) Workspace Monthly Teleconference

Teleconference Information

Date: Thursday, October 20, 2011
Time: 1:00 – 3:00 PM ET
Moderator: Brian Davis

Executive Summary

Dave Billiter shared use cases and requirements from his perspective as Director at The Research Institute at Nationwide Children's Hospital Biopathology Center. We introduced the VCDE WS "Virtual Journal Club" including a presentation by Jyotishman Pathak, author of an abstract "Mapping Clinical Phenotype Data Elements to Standardized Metadata Repositories and Controlled Terminologies: The eMERGE Network Experience" and a presentation by Bastien Rance, author of the article "Integrating clinical research with the Healthcare Enterprise: From the RE-USE project to the EHR4CR platform".

Meeting Materials

-

Attendees

First Name	Last Name	Affiliation
Salvatore	Mungal	Duke University
Hua	Min	George Mason University
Grace	Stafford	The Jackson Laboratory
Rick	Kiefer	Mayo Clinic
Bob	Freimuth	Mayo Clinic
Jyoti	Pathak	Mayo Clinic
Mike	Riben	MD Anderson
John	Dzak	Northwestern University
William	Stephens	OSU
Jonathan	Pierce	OSU
Virginia	Hetrick	Patient Advocate
Carolyn	Peterson	Patient Advocate
Lynne	Wilkens	University of Hawaii
Mukesh	Sharma	Washington University
Mary	McAdams	IMS
Dianne	Reeves	NCI CBIIT
Denise	Warzel	NCI CBIIT
Sherri	De Coronado	NCI CBIIT

Matt	Kennedy	NCI CBIIT
Larry	Wright	NCI CBIIT
Christo	Andonyadis	NCI CBIIT
Lisa	Cole	NCI CBIIT
Dave	Hau	NCI CBIIT
Tommie	Curtis	SAIC
Janice	Chili	SAIC
Brenda	Maeske	SAIC
Mary	Cooper	SAIC
Larry	Brem	SAIC-Frederick
Claire	Wolfe	TerpSys
Riki	Ohira	Booz Allen Hamilton
Mike	Keller	Booz Allen Hamilton
Beth	DiGiulian	Booz Allen Hamilton
Dave	Billiter	Nationwide Children's
Andrew	Buckler	Buckler Biomedical LLC
Don	F	
Bastien	Rance	National Library of Medicine
Bart	Brown	University of Iowa

MEETING NOTES
Discussion –

1. Use Cases, Requirements and Scenarios – Dave Billiter

- Brian Davis attended the ICR and TBPT Workspace Face-to-Face Meeting in August 2011 and heard Dave Billiter speak about their semantic needs that were clear and explicit, he thought it would be worth having Dave present to the VCDE and Architecture Workspaces
- There are several cooperative groups involved that participate in the Group Banking Committee (GBC)
- Biobanking at the Biopathology Center is collecting tissues as well as distributing tissues using a very dynamic workflow
- From Cancer Cooperative Group structure, there are multiple components to a Cooperative Group. There are disease committees that develop and review concepts for study development
    - Within COG there are 11 disease committees
- They are more web services focused and there is a significant amount of semantic interoperability that needs to take place
- Cooperative Groups are moving toward Medidata Rave and this is a significant move towards standardization and interoperability
- There are a lot of legacy operating systems that need to be harmonized and integrated; within cancer cooperative groups, they define data elements as group data elements, which does not map to metadata already defined in the caDSR.

- Brian: There are different users, but this group already had documented who is involved, what needs to get done and what the pain points are for each group and each step of the process
- They are one of two core biobanking services for the The Cancer Genome Atlas (TCGA) program; there is a significant amount of data collected, characterized and shared.
- Q: How are you doing the annotations for the specimen? A: One of the annotation aspects is related to the pathology review. Pathologists review the specimen using forms (Dave's group has built their own web-based digital pathology review system) for electronic submission of annotations.
- Q: Where are the terminologies coming from originally? A: They are from vocabularies, but not all are necessarily in the caDSR (NCIt), but are vocabularies defined by pathology group.
- Data Dictionary is a common term utilized in this community; need to define what a data dictionary is and the definitions can vary based on the different user groups.
- Brian: If there was an easy way to publish data dictionaries that would be a huge win for this group.
- This group is working with Dianne Reeves to review the semantics and stakeholder consensus activities around providing semantics to data dictionaries
- There is a need for significant interaction between biobanking business folks with the informaticians. The mapping aspects of some of the legacy banks can be complicated and they have worked to make the process easier.
- Pain points faced internally include a lot of mapping activities between the different cooperative groups
- The Cooperative Human Tissue Network (CHTN) worked with caDSR team to create metadata for the CHTN tool
- Mapping component is a huge issue at several levels and batch type of functionality to submit "Excel" to provide probabilities on mapping.
- Templates are another challenge, as well as shipping logs (which are often represented as Excel spreadsheets) that could be integrated in mapping tools
- Q: Is there a way that all the paper work can be associated with electronic file? A: Yes, absolutely. Many cancer cooperative groups are further ahead and some are consistently working towards moving to electronic processing of biospecimen. When Virginia was asked to send specimen to a biobank, she received 84 pieces of paper just for that submission.
- Q: If a specific tissue bank is not associated with a cooperative group, is there a mechanism to accept specimen, data, etc.? A: Yes. We also work outside of cancer and can accept other specimen/data.
- Q: Is it possible to bring some of those tooling to an open source development with community code contributed environment, like the CDE mapping, etc.? A: Yes, absolutely. They are very interested in seeing how this can move towards an open source development environment.
- Action Item:

- Brian – See if there is interest to have another meeting with Dave to discuss next steps around follow up on tooling and bringing it to the open source development environment with caBIG®

2. Introducing the VCDE WS "Virtual Journal Club" – Sherri De Coronado, Brian Davis, Riki Ohira

- Link to Virtual Journal Club: https://wiki.nci.nih.gov/x/IQIpAw
- The VCDE Workspace is asking the community to please provide content and ideas to the Virtual Journal Club
- As an example, Sal Mungal suggested Rachel Richesson's article to the group before the VCDE Workspace leadership thought about formally starting this effort
- This is a community-driven effort and we hope participants will participate actively
- Hope to work collaboratively with the Vocabulary Knowledge Center to manage the Virtual Journal Club site, like start listing and voting on articles, etc.
- Action Items:
  - Community – Please start adding suggestions/content to the Virtual Journal Club wiki
  - Community – Send any comments or information to Brian Davis, Sherri De Coronado or Riki Ohira

3. "Mapping Clinical Phenotype Data Elements to Standardized Metadata Repositories and Controlled Terminologies: The eMERGE Network Experience" – JAMIA – Jyotishman Pathak

- Link to abstract: http://jamia.bmj.com/content/18/4/376.short
- Jyoti is presenting the work recently published in JAMIA and is the work of the larger network
- Isaac Kohane wrote that it is critical to access EHR and EMR data to truly leverage informatics
- Through work with the eMERGE data dictionary efforts, they have learned that data aggregation and integration is very difficult if there is not explicit semantics associated with the data
- dbGaP collects data with local terminologies and data standards making it difficult to search for desired data. Part of eMERGE's efforts was to try to standardize the data collected in dbGaP using a simple workflow approach to map and harmonize the metadata.
- Collected Data Dictionaries from 11 sites and did some cleaning to harmonize the variable needs and developed tool that would take variables and list of permissible values that were applicable and executed regular expression string based search.
- There is a video to demonstrate how the eleMap tool works
- Q: Is there a possibility of taking in a spreadsheet from one of Dave's group and doing a mapping of a data dictionary all at once? A: Yes, you can import a spreadsheet. As long as spreadsheet is dbGaP compliant, you can use it to map elements in your data dictionary.

- There is a capability to create your own data dictionary from scratch or using an existing one (allows for drag-and-drop feature to add elements to basket)
- Q: is the link the manuscript in this presentation? A: If you click on the link to abstract, it will take you to the PDF. This was an Editor's choice, so even if you don't have subscription to JAMIA you can still access the article.
- Q: When working with ISO 11179 you went to the permissible value set and stayed away from the object class. Are there other insights how to deal with this issue? A: We struggled early on with dealing with object class, attributes, permissible values, etc. The closest mapping one would be performing and primary interest was in the permissible values. Tried to keep this as simple as possible. Not sure if it is the best approach, but it got the group to a place where they could do some kind of harmonization of data in dbGaP.
- Sal worked with atomic class that could be used throughout entire model and did work with cardiology. Allowed for identification of small set of CDEs based on the atomic class model.
- Action Items:

4. **"Integrating clinical research with the Healthcare Enterprise: From the RE-USE project to the EHR4CR platform" – Journal of Biomedical Semantics – Bastien Rance**

- Link to article: http://www.ncbi.nlm.nih.gov/pubmed/21888989
- Bastien worked specifically on the data integration and mapping aspects and not as much on the design aspects of the paper
- Similar theme for all presentations, this paper focused on enabling integration of EHR data for clinical research
- Action Items.

Next Meeting: Thursday, November 17, 2011; 1:00 – 3:00 PM ET